



PCT/PCT No. 11 MAR 2002

#7

## SEQUENCE LISTING

&lt;110&gt; Strathmann AG &amp; Co.

&lt;120&gt; Virus-Vaccine

&lt;130&gt; P057760

&lt;140&gt;

&lt;141&gt;

&lt;150&gt; 199 07 485.2

&lt;151&gt; 1999-02-12

&lt;160&gt; 12

&lt;170&gt; PatentIn Ver. 2.1

&lt;210&gt; 1

&lt;211&gt; 9709

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 1

tggaagggtct aatttgggtcc caaaaaagac aagagatcct tgatctgtgg atctaccaca 60  
cacaaggcta cttccctgat tggcagaact acacaccagg gccagggatc agatatccac 120  
tgacctttgg atggtgcttc aagttagtac cagttgaacc agagcaagta gaagaggcca 180  
aataaggaga gaagaacagc ttgttacacc ctatgagcca gcatgggatg gaggacccgg 240  
agggagaagt attagtgtgg aagtttgaca gcctcctagc atttcgtcac atggcccgag 300  
agctgcatcc ggagtactac aaagactgct gacatcgagc tttctacaag ggactttccg 360  
ctggggactt tccagggagg tgtggcctgg gcgggactgg ggagtggcga gccctcagat 420  
gctacatata agcagctgct ttttgcctgt actgggtctc tctgggttaga ccagatctga 480  
gcctgggagc tctctggcta actagggaac ccactgctta agcctcaata aagcttgcct 540  
tgagtgtctc aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc 600  
agaccctttt agtcagtgtg gaaaatctct agcagtggcg cccgaacagg gacttgaaag 660  
cgaaagtaaa gccagaggag atctctcgac gcaggactcg gcttgttgaa gcgcgcacgg 720  
caagaggcga ggggcggcga ctggtgagta cgccaaaaat tttgactagc ggaggctaga 780  
aggagagaga tgggtgctgag agcgtcggta ttaagcgggg gagaattaga taaatgggaa 840  
aaaattcggg taaggccagg gggaaagaaa caatataaac taaaacatat agtatgggca 900  
agcagggagc tagaacgatt cgcagttaat cctggccttt tagagacatc agaaggctgt 960  
agacaaatac tgggacagct acaaccatcc cttcagacag gatcagaaga acttagatca 1020  
ttatataata caatagcagt cctctattgt gtgcatcaaa ggatagatgt aaaagacacc 1080  
aaggaagcct tagataagat agaggaagag caaaacaaaa gtaagaaaaa ggcacagcaa 1140  
gcagcagctg acacaggaaa caacagccag gtcagccaaa attaccctat agtgcagaa 1200  
ctccaggggc aaatggtaca tcaggccata tcacctagaa ctttaaatgc atgggtaaaa 1260  
gtagtagaag agaaggcttt cagcccagaa gtaataccca tgttttcagc attatcagaa 1320  
ggagccaccc cacaagattt aaataccatg ctaaaccacag tgggggggaca tcaagcagcc 1380  
atgcaaatgt taaaagagac catcaatgag gaagctgcag aatgggatag attgcatcca 1440  
gtgcatgcag ggcctattgc accaggccag atgagagaa caaggggaag tgacatagca 1500  
ggaactacta gtacccttca ggaacaaata ggatggatga cacataatcc acctatccca 1560  
gtaggagaaa tctataaaa atggataatc ctgggattaa ataaaatagt aagaatgtat 1620  
agccctacca gcattctgga cataagacaa ggaccaaagg aaccctttag agactatgta 1680  
gaccgattct ataaaactct aagagccgag caagcttcac aagaggtaaa aaattggatg 1740  
acagaaacct tgttgggtcca aaatgcgaac ccagattgta agactatttt aaaagcattg 1800  
ggaccaggag cgacactaga agaaatgatg acagcatgtc agggagtggg gggacccggc 1860  
cataaagcaa gagttttggc tgaagcaatg agccaagtaa caaatccagc taccataatg 1920

PCT/PCT No. 11 MAR 2002

atacagaaaag	gcaatttttag	gaaccaaaga	aagactgtta	agtgtttcaa	ttgtggcaaa	1980
gaagggcaca	tagccaaaaa	ttgcagggcc	cctaggaaaa	agggctgttg	gaaatgtgga	2040
aaggaaggac	accaaataag	agattgtact	gagagacagg	ctaatttttt	aggggaagatc	2100
tggccttccc	acaaggggaag	gccaggggaat	tttcttcaga	gcagaccaga	gccaacagcc	2160
ccaccagaag	agagcttcag	gtttgggggaa	gagacaacaa	ctccctctca	gaagcaggag	2220
ccgatagaca	aggaactgta	tccttttagct	tccttcagat	cactcttttg	cagcgacccc	2280
tcgtcacaa	aaagataggg	gggcaattaa	aggaagctct	attagatata	ggagcagatg	2340
atacagtatt	agaagaaatg	aatttgccag	gaagatggaa	acccaaaatg	ataggggggaa	2400
ttggagggtt	tatcaaagta	ggacagtatg	atcagatact	catagaaatc	tgcggacata	2460
aagctatagg	tacagtatta	gtaggaccta	cacctgtcaa	cataattgga	agaaatctgt	2520
tgactcagat	tggctgcact	ttaaattttc	ccatttagtcc	tattgagact	gtaccagtaa	2580
aattaaagcc	aggaatggat	ggcccaaaa	ttaaacaatg	gccattgaca	gaagaaaaaa	2640
taaaagcatt	agtagaaatt	tgtacagaaa	tggaaaagga	agggaaaatt	tcaaaaattg	2700
ggcctgaaaa	tccatacaat	actccagtat	ttgccataaa	gaaaaaagac	agtactaaat	2760
ggagaaaaat	agtagatttc	agagaactta	ataagagaa	tcaagatttc	tgggaagttc	2820
aattaggaat	accacatcct	gcaggggtta	aacagaaaaa	atcagtaaca	gtactggatg	2880
tgggcgatgc	atatttttca	gttcccttag	ataaagactt	caggaagtat	actgcattta	2940
ccatacctag	tataaacaat	gagacaccag	ggattagata	tcagtacaat	gtgcttccac	3000
agggatggaa	aggatcacca	gcaatattcc	agtgtagcat	gacaaaaatc	ttagagcctt	3060
ttagaaaaac	aaatccagac	atagtctctt	atcaatcat	ggatgatttg	tatgtaggat	3120
ctgacttaga	aatatggcag	catagaacaa	aaatagagga	actgagacaa	catctgttga	3180
ggtggggatt	taccacacca	gacaaaaaac	atcagaaaga	acctccattc	ctttggatgg	3240
gttatgaact	ccatcctgat	aaatggacag	tacagcctat	agtgttgcca	gaaaaggaca	3300
gctggactgt	caatgacata	cagaaattag	tgggaaaatt	gaattgggca	agtcagattt	3360
atgcagggat	taaagtaagg	caattatgta	aacttcttag	gggaaccaa	gcactaacag	3420
aagtagtacc	actaacagaa	gaagcagagc	tagaactggc	agaaaacagg	gagattctaa	3480
aagaaccggt	acatggagtg	tattatgacc	catcaaaaga	cttaatagca	gaaatacaga	3540
agcaggggca	aggccaatgg	acatatcaaa	tttatcaaga	gccattttaa	aatctgaaaa	3600
caggaaaata	tgcaagaatg	aagggtgccc	acactaatga	tgtgaaacaa	ttaacagagg	3660
cagtacaaaa	aatagccaca	gaaagcatag	taatattggg	aaagactcct	aaattttaa	3720
taccatata	aaaggaaaca	tgggaagcat	ggtggacaga	gtattggcaa	gccacctgga	3780
ttcctgagt	ggagtttg	aataccctc	ccttagtgaa	gttatggtac	cagtttagaga	3840
aagaacccat	aataggagca	gaaactttct	atgtagatgg	ggcagccaat	agggaaacta	3900
aattaggaaa	agcaggatat	gtaactgaca	gaggaagaca	aaaagttgtc	cccctaacgg	3960
acacaacaaa	tcagaagact	gagttacaag	caattcatct	agctttgcag	gattcgggat	4020
tagaagtaaa	catagtga	gactcacaat	atgcattggg	aatcattcaa	gcacaaccag	4080
ataagagtga	atcagagtta	gtcagtcaaa	taatagagca	gttaataaaa	aaggaaaaag	4140
tctacctggc	atgggtacca	gcacacaaag	gaattggagg	aatgaacaa	gtagatgggt	4200
tggctcagt	tggaatcagg	aaagtactat	ttttagatgg	aatagataag	gccaagaag	4260
aacatgagaa	atatcacagt	aattggagag	caatggctag	tgattttaac	ctaccacctg	4320
tagtagcaaa	agaaatagta	gccagctgtg	ataaatgtca	gctaaaaggg	gaagccatgc	4380
atggacaagt	agactgtagc	ccaggaatat	ggcagctaga	ttgtacacat	ttagaaggaa	4440
aagttatctt	ggtagcagtt	catgtagcca	gtggatatat	agaagcagaa	gtaattccag	4500
cagagacagg	gcaagaaaca	gcatacttcc	tcttaaaatt	agcaggaaga	tggccagtaa	4560
aaacagtaca	tacagacaat	ggcagcaatt	tcaccagtac	tacagttaag	gccgcctgtt	4620
ggtgggcggg	gatcaagcag	gaatttgcca	ttccctacaa	tcccaaaagt	caaggagtaa	4680
tagaatctat	gaataaagaa	ttaaagaaaa	ttataggaca	ggtaagagat	caggctgaac	4740
atcttaagac	agcagtacaa	atggcagtat	tcattccaca	ttttaaaaga	aaagggggga	4800
ttgggggggt	cagtgcagg	gaaagaatag	tagacataat	agcaacagac	atacaaaacta	4860
aagaattaca	aaaacaaatt	acaaaaattc	aaaattttcg	ggtttattac	agggacagca	4920
gagatccagt	ttggaaagga	ccagcaaagc	tcctctggaa	aggtgaaggg	gcagtagtaa	4980
tacaagataa	tagtgacata	aaagttagtc	caagaagaaa	agcaaagatc	atcagggatt	5040
atggaaaaca	gatggcaggt	gatgattgtg	tggcaagtag	acaggatgag	gattaacaca	5100
tggaaaagat	tagtaaaaca	ccatatgtat	atttcaagga	aagctaagga	ctggttttat	5160
agacatcact	atgaaagtac	taatccaaaa	ataagttcag	aagtacacat	cccactaggg	5220
gatgctaaat	tagtaataac	aacatattgg	ggtctgcata	caggagaaag	agactggcat	5280
ttgggtcagg	gagtcctcat	agaatggagg	aaaaagagat	atagcacaca	agtagaccct	5340

gacctagcag	accaactaat	tcatctgcac	tatttttgatt	gtttttcaga	atctgtctata	5400
agaaataacca	tattaggacg	tatagtttagt	cctaggtgtg	aatatcaagc	aggacataac	5460
aaggtaggat	ctctacagta	cttggcacta	gcagcattaa	taaaaccaa	acagataaag	5520
ccacctttgc	ctagtgttag	gaaactgaca	gaggcacagat	ggaacaagcc	ccagaagacc	5580
aagggccaca	gagggagcca	tacaatgaat	ggacactaga	gcttttagag	gaacttaaga	5640
gtgaagctgt	tagacatttt	cctaggatat	ggctccataa	cttaggacaa	catatctatg	5700
aaacttacgg	ggatacttgg	gcaggagtgg	aagccataat	aagaattctg	caacaactgc	5760
tgtttatcca	tttcagaatt	gggtgtcgac	atagcagaat	aggcgttact	cgacagagga	5820
gagcaagaaa	tggagccagt	agatcctaga	ctagagccct	ggaagcatcc	aggaagtcag	5880
cctaaaactg	cttgtagcaa	ttgtctattgt	aaaaagtggt	gctttcattg	ccaagtttgt	5940
ttcatgacaa	aagccttagg	catctcctat	ggcaggaaga	agcggagaca	gcgacgaaga	6000
gctcatcaga	acagtcagac	tcatcaagct	tctctatcaa	agcagtaagt	agtacatgta	6060
atgcaaccta	taatagtagc	aatagtagca	ttagtagtag	caataataat	agcaatagtt	6120
gtgtggtcca	tagtaatcat	agaatatagg	aaaatattaa	gacaaagaaa	aatagacagg	6180
ttaattgata	gactaataga	aagagcagaa	gacagtggca	atgagagtga	aggagaagta	6240
tcagcacttg	tggagatggg	ggtggaaatg	gggcaccatg	ctccttggga	tattgatgat	6300
ctgtagtgt	acagaaaaat	tgtgggtcac	agtcctattat	ggggtaacctg	tgtggaagga	6360
agcaaccacc	actctatttt	gtgcatcaga	tgctaaagca	tatgatcacg	aggtacataa	6420
tgtttgggcc	acacatgcct	gtgtaccac	agaccccaac	ccacaagaag	tagtattggt	6480
aaatgtgaca	gaaaatttta	acatgtggaa	aaatgacatg	gtagaacaga	tgcatgagga	6540
tataatcagt	ttatgggatc	aaagcctaaa	gccatgtgta	aaattaaccc	cactctgtgt	6600
tagtttaaag	tgcactgatt	tgaagaatga	tactaatacc	aatagtagta	gcggggagaat	6660
gataatggag	aaaggagaga	taaaaaactg	ctctttcaat	atcagcacia	gcataagaga	6720
taaggtgcag	aaagaatatg	cattctttta	taaacttgat	atagtaccaa	tagataatac	6780
cagctatagg	ttgataagtt	gtaacacctc	agtcattaca	caggcctgtc	caaaggtatc	6840
ctttgagcca	attcccatac	attattgtgc	ccgggtggt	tttgcgattc	taaaatgtaa	6900
taataagacg	ttcaatggaa	caggaccatg	tacaaatgtc	agcacagtac	aatgtacaca	6960
tggaatcagg	ccagtagtat	caactcaact	gctgttaaat	ggcagtctag	cagaagaaga	7020
tgtagtaatt	agatctgcca	atttcacaga	caatgctaaa	accataatag	tacagctgaa	7080
cacatctgta	gaaattaatt	gtacaagacc	caacaacaat	acaagaaaaa	gtatccgtat	7140
ccagagggga	ccagggagag	catttgttac	aataggaaaa	ataggaaata	tgagacaagc	7200
acattgtaac	attagtagag	caaaatggaa	tgccacttta	aaacagatag	ctagcaaatt	7260
aagagaacaa	tttggaaata	ataaaacaat	aatcttttaag	caatcctcag	gaggggaccc	7320
agaaattgta	acgcacagtt	ttaatttgtg	aggggaattt	ttctactgta	attcaacaca	7380
actgtttaat	agtacttgg	ttaatatgac	ttggagtact	gaagggtcaa	ataacactga	7440
aggaagtgc	acaaatcac	tcccatgcag	aataaaacaa	ttataaaca	tgtggcagga	7500
actggaaaaa	gcaatgtatg	ccccctccat	cagtggacaa	attagatggt	catcaaatat	7560
tactgggctg	ctattaacaa	gagatgggtg	taataacaac	aatgggtccg	agatcttcag	7620
acctggagga	ggcgatatga	gggacaattg	gagaagtga	ttatataaat	ataaagtagt	7680
aaaaattgaa	ccattaggag	tagcaccac	caaggcaag	agaagagtgg	tgacagagaga	7740
aaaaagagca	gtgggaatag	gagctttgtt	ccttgggttc	ttgggagcag	caggaagcac	7800
tatgggctgc	acgtcaatga	cgctgacggt	acaggccaga	caattattgt	ctgatatagt	7860
gcagcagcag	aacaatttgc	tgagggtctat	tgaggcgcaa	cagcatctgt	tgcaactcac	7920
agtctggggc	atcaaacagc	tccaggcaag	aatcctggct	gtggaaagat	acctaagga	7980
tcaacagctc	ctggggattt	ggggttgctc	tggaaaactc	atttgcacca	ctgctgtgcc	8040
ttggaatgct	agttggagta	ataaatctct	ggaacagatt	tggaaataca	tgacctggat	8100
ggagtgggac	agagaaaatta	acaattacac	aagcttaata	cactccttaa	ttgaagaatc	8160
gcaaaaccag	caagaaaaga	atgaacaaga	attattggaa	ttagataaat	gggcaagttt	8220
gtggaattgg	tttaacataa	caaattggct	gtggtatata	aaattattca	taatgatagt	8280
aggaggcttg	gtaggtttaa	gaatagtttt	tgtgtactt	tctatagtga	atagagttag	8340
gcagggatat	tcaccattat	cgtttcagac	ccacctccca	atcccagggg	gacccgacag	8400
gcccgaagga	atagaagaag	aaggtggaga	gagagacaga	gacagatcca	ttcgattagt	8460
gaacggatcc	ttagcactta	tctgggacga	tctgctggagc	ctgtgcctct	tcagctacca	8520
cgcttgaga	gacttactct	tgattgtaac	gaggattgtg	gaacttctgg	gacgcagggg	8580
gtgggaagcg	ctcaaatatt	ggtggaactc	cctacagtat	ttgagtcagg	aactaaagaa	8640
tagtgcctgt	aacttgctca	atgccacagc	catagcagta	gctgagggga	cagataggg	8700
tatagaagta	ttacaagcag	cttatagagc	tattcgccac	atacctagaa	gaataagaca	8760



Ile	Ser	Thr	Ser	Ile	Arg	Asp	Lys	Val	Gln	Lys	Glu	Tyr	Ala	Phe	Phe	
				165					170					175		
Tyr	Lys	Leu	Asp	Ile	Val	Pro	Ile	Asp	Asn	Thr	Ser	Tyr	Arg	Leu	Ile	
			180					185					190			
Ser	Cys	Asn	Thr	Ser	Val	Ile	Thr	Gln	Ala	Cys	Pro	Lys	Val	Ser	Phe	
		195					200					205				
Glu	Pro	Ile	Pro	Ile	His	Tyr	Cys	Ala	Pro	Ala	Gly	Phe	Ala	Ile	Leu	
	210					215					220					
Lys	Cys	Asn	Asn	Lys	Thr	Phe	Asn	Gly	Thr	Gly	Pro	Cys	Thr	Asn	Val	
225					230					235					240	
Ser	Thr	Val	Gln	Cys	Thr	His	Gly	Ile	Arg	Pro	Val	Val	Ser	Thr	Gln	
				245					250					255		
Leu	Leu	Leu	Asn	Gly	Ser	Leu	Ala	Glu	Glu	Asp	Val	Val	Ile	Arg	Ser	
			260					265					270			
Ala	Asn	Phe	Thr	Asp	Asn	Ala	Lys	Thr	Ile	Ile	Val	Gln	Leu	Asn	Thr	
		275					280					285				
Ser	Val	Glu	Ile	Asn	Cys	Thr	Arg	Pro	Asn	Asn	Asn	Thr	Arg	Lys	Ser	
	290					295					300					
Ile	Arg	Ile	Gln	Arg	Gly	Pro	Gly	Arg	Ala	Phe	Val	Thr	Ile	Gly	Lys	
305					310					315					320	
Ile	Gly	Asn	Met	Arg	Gln	Ala	His	Cys	Asn	Ile	Ser	Arg	Ala	Lys	Trp	
				325					330					335		
Asn	Ala	Thr	Leu	Lys	Gln	Ile	Ala	Ser	Lys	Leu	Arg	Glu	Gln	Phe	Gly	
			340					345					350			
Asn	Asn	Lys	Thr	Ile	Ile	Phe	Lys	Gln	Ser	Ser	Gly	Gly	Asp	Pro	Glu	
		355					360					365				
Ile	Val	Thr	His	Ser	Phe	Asn	Cys	Gly	Gly	Glu	Phe	Phe	Tyr	Cys	Asn	
	370					375					380					
Ser	Thr	Gln	Leu	Phe	Asn	Ser	Thr	Trp	Phe	Asn	Ser	Thr	Trp	Ser	Thr	
385					390					395					400	
Glu	Gly	Ser	Asn	Asn	Thr	Glu	Gly	Ser	Asp	Thr	Ile	Thr	Leu	Pro	Cys	
			405						410					415		
Arg	Ile	Lys	Gln	Phe	Ile	Asn	Met	Trp	Gln	Glu	Val	Gly	Lys	Ala	Met	
			420					425					430			
Tyr	Ala	Pro	Pro	Ile	Ser	Gly	Gln	Ile	Arg	Cys	Ser	Ser	Asn	Ile	Thr	
	435						440					445				
Gly	Leu	Leu	Leu	Thr	Arg	Asp	Gly	Gly	Asn	Asn	Asn	Asn	Gly	Ser	Glu	

000100-000000

450					455					460					
Ile 465	Phe	Arg	Pro	Gly	Gly 470	Gly	Asp	Met	Arg	Asp 475	Asn	Trp	Arg	Ser	Glu 480
Leu	Tyr	Lys	Tyr	Lys 485	Val	Val	Lys	Ile	Glu 490	Pro	Leu	Gly	Val	Ala 495	Pro
Thr	Lys	Ala	Lys 500	Arg	Arg	Val	Val	Gln 505	Arg	Glu	Lys	Arg	Ala 510	Val	Gly
Ile	Gly	Ala 515	Leu	Phe	Leu	Gly	Phe 520	Leu	Gly	Ala	Ala	Gly 525	Ser	Thr	Met
Gly	Cys 530	Thr	Ser	Met	Thr	Leu 535	Thr	Val	Gln	Ala	Arg 540	Gln	Leu	Leu	Ser
Asp 545	Ile	Val	Gln	Gln 550	Gln	Asn	Asn	Leu	Leu	Arg 555	Ala	Ile	Glu	Ala	Gln 560
Gln	His	Leu	Leu	Gln 565	Leu	Thr	Val	Trp	Gly 570	Ile	Lys	Gln	Leu	Gln 575	Ala
Arg	Ile	Leu	Ala 580	Val	Glu	Arg	Tyr	Leu 585	Lys	Asp	Gln	Gln	Leu 590	Leu	Gly
Ile	Trp	Gly 595	Cys	Ser	Gly	Lys	Leu 600	Ile	Cys	Thr	Thr	Ala 605	Val	Pro	Trp
Asn 610	Ala	Ser	Trp	Ser	Asn	Lys 615	Ser	Leu	Glu	Gln	Ile 620	Trp	Asn	Asn	Met
Thr 625	Trp	Met	Glu	Trp	Asp 630	Arg	Glu	Ile	Asn	Asn 635	Tyr	Thr	Ser	Leu	Ile 640
His	Ser	Leu	Ile	Glu 645	Glu	Ser	Gln	Asn	Gln 650	Gln	Glu	Lys	Asn	Glu 655	Gln
Glu	Leu	Leu	Glu 660	Leu	Asp	Lys	Trp	Ala 665	Ser	Leu	Trp	Asn	Trp 670	Phe	Asn
Ile	Thr	Asn 675	Trp	Leu	Trp	Tyr	Ile 680	Lys	Leu	Phe	Ile	Met 685	Ile	Val	Gly
Gly 690	Leu	Val	Gly	Leu	Arg	Ile 695	Val	Phe	Ala	Val	Leu 700	Ser	Ile	Val	Asn
Arg 705	Val	Arg	Gln	Gly	Tyr 710	Ser	Pro	Leu	Ser	Phe 715	Gln	Thr	His	Leu	Pro 720
Ile	Pro	Arg	Gly	Pro 725	Asp	Arg	Pro	Glu	Gly 730	Ile	Glu	Glu	Glu	Gly 735	Gly
Glu	Arg	Asp 740	Arg	Asp	Arg	Ser	Ile 745	Arg	Leu	Val	Asn	Gly 750	Ser	Leu	Ala



[illegible][illegible][illegible][illegible][illegible][illegible][illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

[illegible][illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	

[illegible][illegible]



<212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Description of the artificial sequence: synthetic DNA

<220>  
 <221> misc\_feature  
 <222> (3)..(9)  
 <223> BstEII cleavage site

<220>  
 <221> misc\_feature  
 <222> (2143)..(2148)  
 <223> BamHI cleavage site

<400> 9  
 tgggtcaccg tctattatgg ggtgcctgtg tggaaggaag caaccaccac tctattttgt 60  
 gcatcagatg ctaaagcata tgatacagag gtacataatg tttgggccac acatgcctgt 120  
 gtaccacag accccaaccc acaagaagta gtattggtta atgtgacaga aaattttaac 180  
 atgtggaaaa atgacatggg agaacagatg catgaggata taatcagttt atgggatcaa 240  
 agccttaagc catgtgtaaa attaacccca ctctgtgtta gtttaaagtg cactgatttg 300  
 aagaatgata ctaataccaa tagtagtagc gggagaatga taatggagaa aggagagata 360  
 aaaaactgca gcttcaatat cagcacaagc ataagagata aggtgcagaa agaatatgca 420  
 ttcttttata aacttgatat agtaccaata gataatacca gctatagggt gataagttgt 480  
 aacacctcag tgatcacaca ggctgttcca aagggtatcct ttgagccaat tcccatacat 540  
 tattgtgccc cggctgggtt tgcgattcta aaatgtaata ataagacgtt caatggaaca 600  
 ggaccatgta caaatgtcag cacagtacaa tgtacacatg gaattcgacc agtagtatca 660  
 actcaactgc tgttaaattg cagtctagca gaagaagatg tagtaattag atctgccaat 720  
 ttcacagaca atgctaaaac cataatagta cagctgaaca catctgtaga aattaattgt 780  
 acaagaccca acaacaatac aagaaaaagt atccgtatcc agaggggacc agggagagca 840  
 tttgttacia taggaaaaat aggaaatatg agacaagcac attgtaacat ttctagagca 900  
 aaatggaatg ccacttttaa acagatagct agcaaattaa gagaacaatt tggaaataat 960  
 aaaacaataa tctttaagca gtcacccgga ggggacccag aaattgtaac gcacagtttt 1020  
 aattgtggag ggggaatttt ctactgtaat tcaacacaac tgtttaatag tacttggttt 1080  
 aatagtactt ggagtactga aggggtcaaat aacactgaag gaagtgcac aatcacactc 1140  
 ccatgcagaa taaaacaatt tataaacatg tggcaggaag taggaaaagc aatgtatgcc 1200  
 cctcccatca gtggccaaat tagatgttca tcaaataatta ctgggctgct attaaactga 1260  
 gatggtggta ataacaacaa tgggtccgag attttcagac ctggaggagg cgatatgagg 1320  
 gataattgga gaagtgaatt atataaatat aaagtagtaa aaattgaacc attaggagta 1380  
 gcaccacca aggcaaagag acgcgtgggt cagagagaaa agcgcgcagt gggaaatagga 1440  
 gctctgttcc ttgggttctt gggagcagca ggaagcacta tgggctgcag gtcaatgacg 1500  
 ctgacggatc aggcagaca attattgtct gatatagtgc agcagcagaa caatttgctg 1560  
 agggcaattg aggcgcaaca gcatctgttg caactcacag tctggggcat caaacagctc 1620  
 caggcaagaa tcctggctgt ggaaagatac ctaaaggatc aacagctcct ggggatttgg 1680  
 ggttgctctg gaaaactcat ttgcaccact gctgtgcctt ggaatgctag ttggagtaat 1740  
 aaatctctgg aacagatttg gaataacatg acctggatgg agtgggacag agaaattaac 1800  
 aattacacaa gcttaataca ctcttaatt gaagaatcgc aaaaccagca agaaaagaat 1860  
 gaacaagaat tattggaatt agataaatgg gcaagtttgt ggaattgggt taacataaca 1920  
 aattggctgt ggtatataaa attattcata atgatagtag gaggcttggt aggtttaaga 1980  
 atagtttttg ctgtactttc tatagtgaat agagttaggc agggatattc accattatcg 2040  
 tttcagaccc acctcccaat cccgagggga cccgacaggc ccgaaggaat agaagaagaa 2100  
 ggtggagaga gagacagaga cagatccatt cgattagtga acggatcc 2148

<210> 10  
 <211> 6229

“GATC” repeats



ccacacatgc	ctgtgtacc	acagacccca	accacaaga	agtagtattg	gtaaatgtga	1560
cagaaaattt	taacatgtgg	aaaaatgaca	tggtagaaga	gatgcatgag	gatataatca	1620
gtttatggga	tcaaagccta	aagccatgtg	taaaattaac	cccactctgt	gttagtttaa	1680
agtgcactga	tttgaagaat	gatactaata	ccaatagtag	tagcggggaga	atgataatgg	1740
agaaaggaga	gataaaaaac	tgctcttttc	atatcagcac	aagcataaga	gataaggtgc	1800
agaaagaata	tgcattcttt	tataaacttg	atatagtacc	aatagataat	accagctata	1860
ggttgataag	ttgtaacacc	tcagtcatta	cacaggcctg	tccaaaggta	tcctttgagc	1920
caattcccat	acattattgt	gccccggctg	gttttgcgat	tctaaaatgt	aataataaga	1980
cgttcaatgg	aacaggacca	tgtacaaatg	tcagcacagt	acaatgtaca	catggaatca	2040
ggccagtagt	atcaactcaa	ctgctgttaa	atggcagtc	agcagaagaa	gatgtagtaa	2100
ttagatctgc	caatttcaca	gacaatgcta	aaaccataat	agtacagctg	aacacatctg	2160
tagaaattaa	ttgtacaaga	cccaacaaca	atacaagaaa	aagtatccgt	atccagaggg	2220
gaccagggag	agcatttggt	acaataggaa	aaataggaaa	tatgagacaa	gcacattgta	2280
acattagtag	agcaaaatgg	aatgccactt	taaaacagat	agctagcaaa	ttaagagaac	2340
aatttgghaa	taataaaaac	ataatcttta	agcaatcctc	aggaggggac	ccagaaattg	2400
taacgcacag	ttttaattgt	ggaggggaat	ttttctactg	taattcaaca	caactgttta	2460
atagtacttg	gtttaatagt	acttggagta	ctgaagggtc	aaataacact	gaaggaaagt	2520
acacaatcac	actcccatgc	agaataaaac	aatttataaa	catgtggcag	gaagtaggaa	2580
aagcaatgta	tgccctctcc	atcagtggac	aaattagatg	ttcatcaa	attactgggc	2640
tgctattaac	aagagatgg	ggtaataaca	acaatgggtc	cgagatcttc	agacctggag	2700
gaggcgatat	gaggggacaat	tggagaagtg	aattatataa	atataaagta	gtaaaaattg	2760
aaccattagg	agtagcaccc	accaaggcaa	agagaagagt	ggtgcagaga	gaaaaaagag	2820
cagtgggaat	aggagctttg	ttccttgggt	tcttggggagc	agcaggaagc	actatgggct	2880
gcacgtcaat	gacgtgacg	gtacaggcca	gacaattatt	gtctgatata	gtgcagcagc	2940
agaacaattt	gctgagggct	attgaggcgc	aacagcatct	gttgcaactc	acagtctggg	3000
gcatcaaaac	gctccaggca	agaatcctgg	ctgtggaaag	atacctaag	gatcaacagc	3060
tcctggggat	ttgggggtgc	tctggaaaac	tcatttgcac	cactgctgtg	ccttgggaatg	3120
ctagttagg	taataaatct	ctggaacaga	tttggaataa	catgacctgg	atggagtg	3180
acagagaaat	taacaattac	acaagcttaa	tacactcctt	aattgaagaa	tcgcaaaacc	3240
agcaagaaaa	gaatgaacaa	gaattattgg	aattagataa	atgggcaagt	ttgtggaatt	3300
ggtttaacat	aacaaattgg	ctgtggtata	taaaattatt	cataatgata	gtaggaggct	3360
tggtagggtt	aagaatagtt	tttgctgtac	tttctatagt	gaatagagtt	aggcagggat	3420
attcaccatt	atcgtttcag	accacctcc	caatcccag	gggacccgac	aggcccgaag	3480
gaatagaaga	agaaggtgga	gagagagaca	gagacagatc	cattcgatta	gtgaacggat	3540
ccttagcact	tatctgggac	gatctgcgga	gcctgtgcct	cttcagctac	caccgcttga	3600
gagacttact	cttgattgta	acgaggattg	tggaaactct	gggacgcagg	gggtgggaag	3660
ccctcaaata	ttggtgggaat	ctcctacagt	attggagtca	ggaactaaag	aatagtgtctg	3720
ttaacttgct	caatggcaca	gccatagcag	tagctgaggg	gacagatagg	gttatagaag	3780
tattacaagc	agcttataga	gctattcgcc	acatacctag	aagaataaga	cagggccttg	3840
aaaggatttt	gctataagat	gggtggcaag	tggtcaaaaa	gtagtgtgat	tggatggcct	3900
gctgtaaggg	aaagaatgag	acgagctgag	ccagcagcag	atgggggtggg	agcagtatct	3960
cgagatctag	actagaacta	gcttcgatcc	agacatgata	agatacattg	atgagtttgg	4020
acaaaccaca	actagaatgc	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgctat	4080
tgctttat	gtaaccatta	taagctgcaa	taaacaagtt	aacaacaaca	attgcattca	4140
ttttatgttt	caggttcagg	gggaggtgtg	ggaggttttt	taaagcaagt	aaaacctcta	4200
caaatgtggt	atggctgatt	atgatcctgc	ctcgcgcgtt	tcggtgatga	cggtgaaaac	4260
ctctgacaca	tgcagctccc	ggagacgggt	acagcttgct	tgtaagcggg	tgccggggagc	4320
agacaagccc	gtcagggcgc	gtcagcgggt	gttggcgggt	gtcggggcgc	agccatgacc	4380
cagtcacgta	gcgatagcgg	agtgtatact	ggcttaacta	tgccggcatca	gagcagattg	4440
tactgagagt	gcaccatatt	tcgggcccgcg	ttgctggcgt	ttttccatca	gtccgcctcc	4500
cctgacgagc	atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaacc	gacaggacta	4560
taaagatacc	aggcgtttcc	ccctggaagc	tcctctgtgc	gctctcctgt	tcggacctgt	4620
ccgcttaccg	gatacctgtc	cgcctttctc	ccttcgggaa	gcgtggcgct	ttctcatagc	4680
tcacgtctga	ggtatctcag	ttcgggtgag	gtcgttcgct	ccaagctggg	ctgtgtgcac	4740
gaacccccg	ttcagcccga	ccgctgcgcc	ttatccggta	actatcgtct	tgagtccaac	4800
ccggtaaagc	acgacttatc	gccactggca	gcagccactg	gtaacaggat	tagcagagcg	4860
aggatatgtag	gcgggtgctac	agagttcttg	aagtgggtggc	ctaactacgg	ctacactaga	4920

```

aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 4980
agctcttgat cccgcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag 5040
cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgggggtct 5100
gacgctcagt ggaacgaaaa ctcacgttaa gggatttttg tcatgagatt atcaaaaagg 5160
atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat 5220
gagtaaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc 5280
tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg 5340
gagggccttac catctggccc cagtgcctga atgataccgc gagaccacg ctcaccggct 5400
ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggctcctgca 5460
actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg 5520
ccagttaata gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg 5580
tcgtttggta tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc 5640
cccattgtgt gcaaaaaagc ggtagctcc ttcggctctc cgatcgttgt cagaagtaag 5700
ttggcccgag tgttatcact catgggtatg gcagcactgc ataattctct tactgtcatg 5760
ccatccgtaa gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag 5820
tgtatgccc gaccgagttg ctcttgccc ggcgcaatac gggataatac cgcgccacat 5880
agcagaactt taaaagtgt catcattgga aaacgttctt cggggcgaaa actctcaagg 5940
atcttaccgc tgttgagatc cagttcgatg taaccactc gtgcacccaa ctgatcttca 6000
gcatctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgcccga 6060
aaaaagggaa taagggcgac acggaaatgt tgaatactca tactcttctt ttttcaatat 6120
tattgaagca tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag 6180
aaaaataaac aaataggggt tccgcgcaca tttccccgaa aagtgccac 6229

```

```

<210> 11
<211> 860
<212> DNA
<213> Human immunodeficiency virus

<220>
<221> misc_feature
<222> (1)..(860)
<223> PI-932 original sequence V1-V2-V3-loop

```

```

<400> 11
tgtgtaccca cagaccccaa cccacaaaag gtagtattgg aaaatgtgac agaaaatttt 60
aacatgtgga aaaatgacat ggtagaacag atgcatgagg atataatcaa tttatgggat 120
caaagcctaa agccatgtgt aaaactaacc ccactctgtg ttacttttaa ttgcactgat 180
gctgatttaa attgcaataa tactgattta aattgcacta aagctaattt ggggaaaaat 240
actcataaca atactattag tgggaaaata atagagaaaag tagaaataaa aaactgctct 300
ttcaagggtca ccacaggcat aagggataag atgcaaaaag aatatgcact tttgaataaa 360
cttgatatag taccaataga taatgataag aataatacta actttatatt gataagttgt 420
aacacctcga ccattacaca ggctgttcca aaggtatcct ttgagccaat tcccatacat 480
ttttgtgccc cggctggttt tgcgattcta aagtgtaatg aaaagagtta cagtggaaaa 540
ggaccatgta aaaatgtcag cacagtacaa tgtacacatg gaattaggcc agtagtgtca 600
actcaactgc tgttgaatgg cagtctagca gaaaaagaag tagtaattag atctgagaat 660
ttcacagaca atgctaaaac cataatagta cagctgaagg aatctgtaaa cattacttgt 720
ataagacccc acaacactgt aacagacagg atacatatag ggccagggag atcatttcat 780
acaacaagaa aaataaaagg agatataaga caagcacatt gtagccttag gagaaaagat 840
tggataaaca ctttacaaga                                     860

```

```

<210> 12
<211> 870
<212> DNA
<213> Artificial Sequence

```

[illegible]

<400> 12

tgtgtacc	cagaccccaa	cccacaaaag	gtagtattgg	aaaatgtgac	agaaaaat	60
aacatgtgga	aaaatgacat	ggtagaacag	atgcatgagg	atataatcaa	tttatgggat	120
caaagcctta	agccatgtgt	aaaactaacc	ccactctgtg	ttacttttaa	ttgcactgat	180
gctgatttaa	attgcaataa	tactgattta	aattgcacta	aagctaattt	ggggaaaaat	240
actcataact	gcagtattag	tgggaaaata	atagagaaag	tagaaataaa	aaactgctct	300
ttcaagggtca	ccacaggcat	aagggataag	atgcaaaaag	aatatgcact	tttgaataaa	360
cttgatatag	taccaataga	taatgataag	aataatacta	actttatatt	gataagttgt	420
aacacctcgg	tgatcacaca	ggcctgtcca	aaggatatcct	ttgagccaat	tcccatacat	480
ttttgtgcc	cggctggttt	tgcgattcta	aagtgtaatg	aaaagagtta	cagtggaaaa	540
ggaccatgta	aaaatgtcag	cacagtacaa	tgtacacatg	gaattcggcc	agtagtgtca	600
actcaactgc	tgttgaatgg	cagtctagca	gaaaaagaag	tagtaattag	atctgagaat	660
ttcacagaca	atgctaaaac	cataatagta	cagctgaatg	aatctgtaaa	cattacttgt	720
ataagacccc	acaacactgt	acaacagcag	atacatagag	ggccagggag	atcatttcat	780
acaacaagaa	aaataaaaagg	agatataaga	caagcacatt	gtagcctttc	tagaaaagat	840
tggaataaca	ctttacaaga	gatagctagc				870